

were used. The sequences of contigs generated by the OSS strategy in AutoAssemblerTM (PE Applied Biosystems, Foster City, CA; Cat. No: 903227) and the gap-filling sequencing trace files were imported into SequencherTM (Gene Codes Corp., Ann Arbor, MI) for overlapping and editing. The sequences generated by the total shotgun strategy were assembled using Phred and Phrap and edited using Consed (<http://chimera.biotech.washington.edu/uwgc/projects.htm>) and GFP (Genome Reconstruction Manager for Phrap), version 1.2 (<http://stork.cellb.bcm.tmc.edu/gfp/>).

PCR-Based gap filling Strategy:

Primers were designed based on the 5'- and 3'-end sequenced of each contig, avoiding repetitive and low quality sequence regions. All primers were designed to be 19-24-mers with 50-70% G/C content. Oligos were synthesized and gel-purified by standard methods.

Since the orientation and order of the contigs were unknown, permutations of the primers were used in the amplification reactions. Two PCR kits were used: first, XL PCR kit (Perkin Elmer, Norwalk, CT; Cat. No.: N8080205), with extension times of approximately 10 minutes; and second, the Taq polymerase PCR kit (Qiagen Inc., Valencia, CA; Cat. No.: 201223) was used under high stringency conditions if smeared or multiple products were observed with the XL PCR kit. The main PCR product from each successful reaction was extracted from a 0.9% low melting agarose gel and purified with the Geneclean DNA Purification kit prior to sequencing.

Analysis:

The identification and characterization of coding regions was carried out as follows: First, repetitive sequences were masked using RepeatMasker (A.F.A. Smit & P. Green, http://ftp.genome.washington.edu/RM/RM_details.html) which screens DNA sequences in FastA format against a library of repetitive elements and returns a masked query sequence. Repeats not masked were identified by comparing the sequence to the GenBank database using WUBLAST2.0 [Altschul, S & Gish, W., Methods Enzymol. 266: 460-480 (1996); <http://blast.wustl.edu/blast/README.html>] and were masked manually.

Next, known genes were revealed by comparing the genomic regions against Genentech's protein database using the WUBLAST2.0 algorithm and then annotated by aligning the genomic and cDNA sequences for each gene, respectively, using a Needleman-Wunch (Needleman and Wunsch, J. Mol. Biol. 48: 443-453 (1970) algorithm to find regions of local identity between sequences. The strategy results in detection of all exons of the five known genes in the region, THPO, TRAP2, eIF4g, CLCN2 and hRPB17 (see below).

Known genes

Map position

eukaryotic translation initiation factor 4 gamma

3q27-qter

thrombopoietin

3q26-q27

chloride channel 2

3q26-qter

TNF receptor associated protein 2

not previously mapped

RNA polymerase II subunit hRPB17

not previously mapped

Finally, novel transcription units were predicted using a number of approaches. CpG islands (S. Cross & Bird, A., Curr. Opin. Genet. Dev. 5: 109-314 (1995) islands were used to define promoter regions and were

identified as clusters of sites cleaved by enzymes recognizing GC-rich, 6 or 8-mer palindromic sequences (NotI, NarI, BssHII, XhoI. CpG islands are usually associated with promoter regions of genes. WUBLAST2.0 analysis of short genomic regions (10-20 kb) versus GenBank revealed matches to ESTs. The individual EST sequences (or where possible, their sequence chromatogram files) were retrieved and assembled with Sequencer to provide a theoretical cDNA sequence (DNA36443). GRAIL2 (ApoCom Inc., Knoxville, TN, command line version for the DEC alpha) was used to predict a novel exon. The five known genes in the region served as internal controls for the success of the GRAIL algorithm.

Isolation:

A partial endothelin converting enzyme-2 (ECE-2) cDNA clone was isolated by first splicing in silico the ECE-2 exons predicted in the genomic sequence to generate a putative sequence (DNA36443). An oligonucleotide probe: GAAGCAGTGCAGCCAGCAGTAGAGAGGGCACCTGCTAAGA) (SEQ ID NO:530) was designed and used to screen a human fetal small intestine library (LIB110) and internal PCR primers (36443f1) (ECE2.f:ACGCAGCTGGAGCTGGTCTTAGCA) (SEQ ID NO:531) and (36443r1) (ECE2.r) (GGTACTGGACCCCTAGGGCCACAA) (SEQ ID NO:532) were used to confirm clones hybridizing to the probe prior to sequencing. One positive clone was obtained, however this cDNA (DNA49830) represented a partially spliced transcript containing appropriately spliced exons 1 through 6, followed by intron 6 sequence. The oligo dT primer annealed to a polyA-stretch within an Alu element present in intron 6. An additional ECE-2 cDNA fragment (DNA49831) was obtained by PCR from a human fetal kidney library (LIB227) with primers designed from the presumed cDNA sequence [36443f3: CCTCCCAGCCGAGACCAAGTGG (SEQ ID NO:533) and 36443r2: GGTCTATAAGGGCCAAGACC (SEQ ID NO:534)]. This PCR product extended from exon 13 into the 3' untranslated region in exon 18.

A full length endothelin converting enzyme 2 (ECE-2) cDNA clone (DNA55800-1263) was isolated from an oligo-dT-primed human fetal brain library. RNA from human fetal brain tissue (20 weeks gestation, #283005)(SRC175) was isolated by guanidine thiocyanate and 5 µg used to generate double stranded cDNA which was cloned into the vector pRK5E. The 3' -primer (pGACTAGTTCTAGATCGCGAGCGGCCCTTTTTTTTTTTTTTTT) (SEQ ID NO:535) and the 5 -linker (pCGGACGCGTGGGTCGA) (SEQ ID NO:536) were designed to introduce XhoI and NotI restriction sites. The library was screened with PCR primers [36443pcrf1: CGGCCGTGATGGCTGGTGACG (SEQ ID NO:537) and 36443r3: GGCAGACTCCTTCCTATGGG (SEQ ID NO:538)] designed from the partial human ECE-2 cDNA sequences (DNA49830 and DNA49831). PCR products were cloned into the vector pCR2.1-TOPO (Invitrogen Corp., Carlsbad, CA, Cat. No. K4500-01) and sequenced with DYE-terminator chemistry as described above.

EXAMPLE 98: Northern Blot and in situ RNA Hybridization Analysis for PRO403

Expression of PRO403 mRNA in human tissues was examined by Northern blot analysis. Human polyA+ RNA blots derived from human fetal and adult tissues (Clontech, Palo Alto, CA; Cat. Nos. 7760-1, 7756-1 and 7755-1) were hybridized to a [32P-α]dATP-labelled cDNA fragments from probe based on the full length PRO403 cDNA. Blots were incubated with the probes in hybridization buffer (5X SSPE; 2X Denhardt's solution; 100 mg/mL denatured sheared salmon sperm DNA; 50% formamide; 2% SDS) for 18 hours at 42°C,

washed to high stringency (0.1XSSC, 0.1% SDS, 50°C) and autoradiographed. The blots were developed after overnight exposure by phosphorimager analysis (Fuji).

PRO403 mRNA transcripts were detected. Analysis of the expression pattern showed the strongest signal of the expected 3.3 kb transcript in adult brain (highest in the cerebellum, putamen, medulla, and temporal lobe, and lower in the cerebral cortex, occipital lobe and frontal lobe), spinal cord, lung and pancreas and higher levels of a 4.5 kb transcript in fetal brain and kidney.

EXAMPLE 99: Use of PRO Polypeptide-Encoding Nucleic Acid as Hybridization Probes

The following method describes use of a nucleotide sequence encoding a PRO polypeptide as a hybridization probe.

DNA comprising the coding sequence of of a PRO polypeptide of interest as disclosed herein may be employed as a probe or used as a basis from which to prepare probes to screen for homologous DNAs (such as those encoding naturally-occurring variants of the PRO polypeptide) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO polypeptide-encoding nucleic acid-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 100: Expression of PRO Polypeptides in *E. coli*

This example illustrates preparation of an unglycosylated form of a desired PRO polypeptide by recombinant expression in *E. coli*.

The DNA sequence encoding the desired PRO polypeptide is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the specific PRO polypeptide coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.